Genes for \(\alpha\)-arabinofuranosyl and ribitol catabolism from *Klebsiella pneumoniae*

H. Heuel, A. Shakeri-Garakani, S. Turgut and J. W. Lengeler

Author for correspondence: J. W. Lengeler. Tel: +49 541 969 2290. Fax: +49 541 969 2293.

e-mail: lengeler@sfbbio1.biologie.uni-osnabrueck.de

The enzymes for catabolism of the pentitols \(\alpha\)-arabinofuranosyl (Dal) and ribitol (Rbt) and the corresponding genes from *Klebsiella pneumoniae* (dal and rbt) and *Escherichia coli* (att and rtl) have been used intensively in experimental evolutionary studies. Four dal and four rbt genes from the chromosome of *K. pneumoniae* 1033-S514 were cloned and sequenced. These genes are clustered in two adjacent but divergently transcribed operons and separated by two convergently transcribed repressor genes, dalR and rbtR. Each operon encodes an NAD-dependent pentose dehydrogenase (dalD and rbtD), an ATP-dependent pentulose kinase (dalK and rbtK) and a pentose-specific ion symporter (dalT and rbtT). Although the biochemical reactions which they catalyse are highly similar, the enzymes showed interesting deviations. Thus, DalR (313 aa) and RbtR (270 aa) belong to different repressor families, and DalD (455 aa) and RbtD (248 aa), which are active as a monomer or as tetramers, respectively, belong to different dehydrogenase families. Of the two kinases (19.3% identity), DalK (487 aa) belongs to the subfamily of short \(\alpha\)-xylulokinases and RbtK (\(\alpha\)-ribulokinase; 535 aa) to the subfamily of long kinases. The repressor, dehydrogenase and kinase genes did not show extensive similarity beyond local motifs. This contrasts with the ion symporters (86.6% identity) and their genes (82.7% identity). Due to their unusually high similarity, parts of dalT and rbtT have previously been claimed erroneously to correspond to 'inverted repeats' and possible remnants of a 'metabolic transposon' comprising the dal and rbt genes. Other characteristic structures, e.g. a secondary att\(L\) site and chi-like sites, as well as the conservation of this gene group in *E. coli* C are also discussed.

**Keywords:** \(\alpha\)-arabinofuranosyl, ribitol, dal, rbt, *Klebsiella pneumoniae*, enteric bacteria

**INTRODUCTION**

The pentitols ribitol (adonitol; Rbt) and \(\alpha\)-arabinofuranosyl (\(\alpha\)-arabinitol; Dal) are used by 85% of *Klebsiella* strains and by 10% of *Escherichia coli* strains, e.g. *E. coli* C, but not *E. coli* K-12 and B (Reiner, 1975). Their metabolism in both organisms comprises similar catabolic steps but separate and inducible genes and enzymes. Metabolism starts with an \(H^+\) symporter for transport and proceeds as summarized in Fig. 1. The corresponding genes are named dal and rbt in *Klebsiella*, but att and rtl in *E. coli*. They are controlled by repressors DalR and RbtR and the inducers \(\alpha\)-arabinofuranosyl and \(\alpha\)-ribulose, respectively (references in Hartley, 1984b; Mortlock, 1984).

The genetics of pentitol catabolism in enteric bacteria is of interest for several reasons. (i) The *dal* and *rbt* genes are adjacent and oppositely oriented, perhaps indicating a gene duplication and inversion which occurred during their evolution (Charnetzky & Mortlock, 1974c; Neuberger & Hartley, 1979; Scangos & Reiner, 1978). (ii) The *dal* and *rbt* genes are flanked by long (about 1400 bp) inverted repeats and are perhaps part of a 'metabolic transposon' (Link & Reiner, 1982). (iii) The *dal* and *rbt* genes are replaced in *Dal* \(Rbt^+\) strains by the non-homologous *gat* genes for galactitol metabolism (genotypic exclusion), i.e. these genes map at identical positions but appear to be mutually exclusive (Link & Reiner, 1983; Reiner, 1975; Woodward & Charles, 1983). (iv) Pentitol catabolism has been used extensively (for a review see Mortlock, 1982) as a model for studying the acquisition of new metabolic capacities, in particular...
experimental evolution in chemostats involving gene duplications and amplifications (Hartley, 1984a; Mortlock, 1984; Wu et al., 1968).

Despite their importance in evolutionary studies, for most 

\( \text{d} \) and \( \text{r} \) genes only partial sequences are available; neither their exact number nor their nature are known, nor have comparisons for all genes and gene products been presented. Furthermore, contradictory results concerning gene location within the operons of \( \text{Klebsiella pneumoniae} \) have been published (Charnetzky & Mortlock, 1974c; Wu et al., 1985). Different nomenclature for the genes of different \( \text{Klebsiella} \) strains has added to the confusion. In this study we present the complete DNA sequence of the 

\( \text{d} \) and \( \text{r} \) genes from \( \text{K. pneumoniae} \) and describe their location within two operons. We also show that the \( \text{d} \) and \( \text{r} \) genes, although encoding enzymes catalysing similar enzymic reactions, are non-homologous, except for the transport systems, and are not flanked by long inverted repeats. Finally, we propose an amended genetic nomenclature for both systems. 

\( \text{r} \) and \( \text{d} \) (Charnetzky & Mortlock, 1974c) are renamed \( \text{r} \) and \( \text{d} \) (for repressor), \( \text{rb} \) and \( \text{db} \) (Charnetzky & Mortlock, 1974c) are renamed \( \text{rb} \) and \( \text{db} \) (for promoter) and \( \text{r} \) and \( \text{a} \) (Reiner, 1975) now correspond to \( \text{a} \) and \( \text{a} \) to comply with official genetic nomenclature.

**METHODS**

**Bacterial strains, plasmids and culture conditions.** Strain KAY2026 (Sprenger & Lengeler 1984) used in this study as the source of the \( \text{d} \) and \( \text{r} \) genes is a phage P1-sensitive Arg^Gua^ derivative of \( \text{Klebsiella pneumoniae} \) (formerly \( \text{Aerobacter aerogenes} \) 1033-5P14 (Tanaka et al., 1967). \( \text{Escherichia coli} \) (at \( \text{r} \)), HB101, mutants BL21(dED3) and JM109 of \( \text{E. coli} \) K-12, and plasmids pSU18, pBluescript II SK(+) (Ausubel et al., 1990), and pHEX3 and pHEX5 (Heuel et al., 1997) have all been described. Minimal and complex media, and MacConkey indicator plates containing 1% carbon source have been described previously (Lengeler, 1973).

**Transport and enzyme tests.** Transport tests using \( ^{14} \text{C}- \)labelled \( \text{d} \)-arabinol and \( \text{r} \)-mannitol and ribitol were done as described previously taking samples between 0 and 60 s (Heuel et al., 1997). \( \text{d} \)-Arabinol and ribitol dehydrogenases were tested in cell extracts also as described previously (Lengeler & Lin, 1972). To test the inducibility of the \( \text{d} \) and \( \text{r} \) genes, cells were grown exponentially in minimal medium with glycerol as carbon source. For induction, \( \text{d} \)-arabinol or ribitol (0.2%) were added and the cells were harvested after one further generation. They were washed twice in minimal medium before being tested.

**Isolation of plasmid DNA, restriction analysis and cloning procedures.** All manipulations with recombinant DNA were carried out by standard procedures as described by Ausubel et al. (1990). Preparation of plasmid DNA was done by phenol extraction (Sambrook et al., 1989) or by using the JetStar DNA purification system (Genomed), while the isolation of chromosomal DNA from KAY2026 was according to Neumann et al. (1992). Restriction enzymes from various commercial sources were used according to the recommendations of the suppliers. Oligonucleotides for sequencing or PCR were purchased either from TIB Molbiol Syntheselabor or from Life Technologies. DNA amplification by PCR was done according to Saiki et al. (1988) with Tag polymerase from Boehringer Mannheim and an Air Thermo-Cycler 1605 from Idaho Technology.

**DNA sequencing and sequence analysis.** To sequence the \( \text{d} \) genes cloned into pHHL104 and the \( \text{r} \) genes cloned into pFCK1 (Fig. 2), known restriction sites were used to obtain smaller clones, the ends of which could be sequenced after subcloning. Where needed, longer fragments were shortened by exonuclease III deletion. Finally, specific DNA primers were used to complete the sequences flanking restriction sites. All DNA sequences were determined on both strands by the dyeoxy chain-termination method with the T7 sequencing kit from Pharmacia. Computer analysis was done with the GenMon 4.3 program from the Gesellschaft für Biotechnologische Forschung (Braunschweig, Germany) and the BLAST programs and database services provided by the National Center for Biotechnology Information (Bethesda, MD, USA).

**T7 RNA polymerase-dependent overexpression and primer extension analysis.** T7 RNA polymerase-dependent overexpression of the \( \text{d} \) gene products and of \( \text{r} \) was done according to a protocol described previously (Nobelman & Lengeler, 1996) using derivatives of medium-copy-number plasmids pHX3 and pHX5 (Heuel et al., 1997) and strain BL21(dED3). The original high-copy-number expression vectors of Tabor & Richardson (1985) or pBluescript II SK(+) (Ausubel et al., 1990) were exceedingly unstable and could not be used. Primer extension experiments were performed according to Ausubel et al. (1990). To determine the transcription initiation sites of \( \text{d} \) and \( \text{r} \), total cellular RNAs were prepared from HB101(pPHHL101) and from HB101(pPHHL102) (\( \text{d} \)). The cells were inoculated into 50 ml of 2 X TY plus chloramphenicol (25 \( \mu \)g ml\(^{-1}\)) to an OD\(_{600}\) of 0.1 and harvested when the cultures reached an...
OD_{660} of 0.5. After centrifugation, the cells were washed, resuspended in 400 µl RNA buffer, lysed and treated as described by Ausubel et al. (1990) to isolate total RNAs. Primer extension reactions were also performed according to Ausubel et al. (1990) using a 31 bp primer (see Fig. 4) for both strands. Primers (50–100 ng) were separated with [γ-32P]ATP and T4 polynucleotide kinase. To remove labelled ATP, a simplified restriction map and essential clones discussed in the text are also shown. The dal genes were cloned in the direction of the lacZ promoter of the vector in pHHL101 but in the opposite direction in pHHL103. Circed restriction sites are conserved between Klebsiella and E. coli. C, BamHI; C, Clal; H, HindIII; K, KpnI; N, NdeI; P, PstI; S, Smal; Sp, Sphi; X, Xhol.

RESULTS AND DISCUSSION

Cloning of the dal and rbt genes from the chromosome of K. pneumoniae

Previous studies with genes D and RNA from two different strains of Klebsiella (Charnerzsky & Mortlock, 1974a, b, c; Hartley, 1984b) and from E. coli C (Link & Reiner, 1983) indicated that the dal and rbt (atl and rlt) genes were contiguous. Hartley and co-workers originally transferred the genes from K. pneumoniae 1033-3P14, by means of P1-transduction, into the chromosome of E. coli K-12 where the genes inserted at a similar place (near 46 min) as in E. coli C (Rigby et al., 1976). Using a secondary att site near the pentitol-specific genes and a phage λ derivative, Neuberger & Hartley (1979) cloned the dal and rbt genes onto phage λ p’ rbt dal. The phage contained rbtDK and dalDK on a 10-9 kbp insert as identified by growth and enzyme tests; it probably also contained rbtR and dalR because the clones conferred control properties (inducibility) identical to those of parental strain 1033-3P14. The gene order was reported to be dalKD-dalRA-rbtR-rbtDK (Fig. 2). This contradicts the order reported by Charnerzsky & Mortlock (1974c) as given as dalKD-rbtB-dalB-rbtDK. It is, however, equivalent to the atl rlt gene order as determined by P1-transduction for E. coli C (Scangos & Reiner, 1978).

Because these discrepancies could be due to chromosomal rearrangements that might have occurred during cloning into phage λ and because neither all dal (atl) nor all rbt (rlt) genes encoding d-arabinol and ribitol transport and catabolism had been cloned and identified, we wanted to reclone all genes and map them completely. Hence, chromosomal DNA was isolated from K. pneumoniae KAY2026 (Sprenger & Lengeler, 1984), one of the strains used by Lin, Hartley and others in their evolutionary studies (references in Mortlock, 1982). The DNA was treated with HindIII before being cloned into the medium-copy-number plasmid pSU18 and transformed into E. coli HB101. This mtlA1 mutant lacks the mannitol-specific phosphotransferase system (IMt) as well as the atl and rlt genes and hence the capacity to grow on d-mannitol, d-arabinol and ribitol. Transformants were first plated on MacConkey indicator plates containing d-arabinol or d-mannitol. IMt negative strains can be suppressed to a Dal' Mtl' phenotype provided a DalT transporter (dalT or atlT), which also transports free d-mannitol, and a d-arabinol dehydrogenase (dalD or aldD), which also converts d-mannitol to d-fructose, are expressed constitutively (Tanaka et al., 1967; Lengeler, 1975; Aukemeyer et al., 1991; Heuel et al., 1997). Among several thousand transformants containing HindIII DNA fragments, one Dal' (strong) Mtl' (weak) but Rtl' colony was found. dalD and dalD were subcloned and located on a 473 kbp fragment. The fragment was reisolated and religated into the polylinker site of pSU18. The subclones were transformed into strain JM109 and the transformants plated on LB-Cam plates (Lengeler, 1975), which were supplemented with 1 mM IPTG and 0.02% X-Gal. From white colonies, plasmids pHHL101 and pHHL103 (Fig. 2) were obtained. Based on their restriction maps, which could be compared to the known maps of dal genes (Hartley, 1984b; Wu et al., 1985; Heuel et al., 1997), it could be concluded that on both plasmids dalD, dalK (encoding a D-xylulose kinase) and dalD were cloned, but in the opposite direction. After transformation into HB101, all transformants from both plasmids had the same Dal' Mtl' phenotype. This result implies that the dal genes are expressed constitutively in both clones, i.e. that no DalR-dependent repression occurs. Because in pHHL103 the dal

---

Fig. 2. Structure of the dal and rbt operons from K. pneumoniae. The operons with genes D (dehydrogenase), K (kinase), T (transporter) and R (repressor) were cloned from the chromosome of strain 1033-3P14. The genes are arranged and transcribed (direction of arrows) as indicated. They are flanked by dhnA, also found in E. coli K-12, and an ORF, orfT, of unknown function not found in E. coli. dhnA, orfT and rbtRD sequences by Hartley (1984b) are indicated by broken arrows. The secondary att site, the area defined as inverted repeats (IR), chi-like recombination sites (chi), the G+C content (mol %) of the ORFs, a simplified restriction map and essential clones discussed in the text are also shown. The dal genes were cloned in the direction of the lacZ promoter of the vector in pHHL101 but in the opposite direction in pHHL103. The phage A derivative, Neuberger & Hartley (1979) cloned the dal genes onto phage λ p’ rbt dal. The phage contained rbtDK and dalDK on a 10-9 kbp insert as identified by growth and enzyme tests; it probably also contained rbtR and dalR because the clones conferred control properties (inducibility) identical to those of parental strain 1033-3P14. The gene order was reported to be dalKD-dalRA-rbtR-rbtDK (Fig. 2). This contradicts the order reported by Charnerzsky & Mortlock (1974c) as given as dalKD-rbtB-dalB-rbtDK. It is, however, equivalent to the atl rlt gene order as determined by P1-transduction for E. coli C (Scangos & Reiner, 1978).
H. HEUEL and OTHERS

+ Purified monomer and dimer in parentheses.

4-7 kb HindIII fragment of pHHL101 was isolated next and cloned into the HindIII site of pHHL102. The new plasmid, pHLH104, carried dalRDKT on a 5.9 kb HindIII–BamHI fragment (Fig. 2). The correct insertion of the dal genes on pHHL104 was ascertained by restriction analysis using NdeI, SphI and XhoI, three restriction enzymes with single, asymmetrically located cutting sites within the dal genes. Cells of HB101 transformed with pHHL104 exhibited a Dal+ (strong) phenotype but remained Mtl-. This is the expected phenotype for a dalR+ strain because D-mannitol is not an inducer for the D-arabinitol-specific DalR repressor. The restriction map of xp rbt dal (Hartley, 1984b) corresponds almost exactly to our map of the 10.93 kbp HindIII–BamHI fragment covering the rbt and dal genes (Fig. 2). Because we isolated the fragment directly from the chromosome of the Klebsiella strain used by these authors, this map should correspond to the genuine map.

**Sequencing of the dal and rbt genes from K. pneumoniae and identification of their gene products**

We sequenced the entire 10.93 kbp fragment covering the region from orf1 to dhnA completely on both strands, except for rbtR and rbtD. First, the sequence data revealed the new genes dalT and rbtT encoding a D-arabinitol- and a ribitol-specific transporter (Heuel et al., 1997), respectively, and second, proved the gene order to be as proposed by Hartley (1984b) and as summarized in Fig. 2. In particular, the dal and rbt genes are adjacent, arranged in a symmetrical but divergent way. The operons are separated by a control region (2.2 kbp) encoding the DalR and RbtR repressors. Because dalR and dalDKT are transcribed from opposite strands, and correspondingly for the rbt genes, dalR and rbtR are transcribed convergently as indicated.

### Table 1. Gene products of the dal and rbt operons

<table>
<thead>
<tr>
<th>Gene product</th>
<th>Function</th>
<th>No. of aa</th>
<th>Observed molecular mass (kDa)</th>
<th>Calculated molecular mass (Da)</th>
<th>Inducer or major substrate</th>
<th>Type or family</th>
<th>Identical residues (%)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>DalR</td>
<td>Repressor</td>
<td>313</td>
<td>34.9†</td>
<td>34.940</td>
<td>d-Arabinitol</td>
<td>DeoR</td>
<td>None</td>
</tr>
<tr>
<td>RbtR</td>
<td></td>
<td>270</td>
<td>–</td>
<td>29.700</td>
<td>d-Ribulose</td>
<td>Lac/GalR</td>
<td>None</td>
</tr>
<tr>
<td>DalD</td>
<td>NAD dehydrogenase</td>
<td>455</td>
<td>46.5†</td>
<td>50.050</td>
<td>d-Arabinitol</td>
<td>Long (monomer)‖</td>
<td>None</td>
</tr>
<tr>
<td>RbtD</td>
<td></td>
<td>248</td>
<td>27.0†</td>
<td>27.280</td>
<td>Ribitol</td>
<td>Short (tetramer)‖</td>
<td>19.3</td>
</tr>
<tr>
<td>DalK</td>
<td>ATP kinase</td>
<td>487</td>
<td>54.0 (110);‡</td>
<td>53.370</td>
<td>d-Xylulose</td>
<td>Hexokinase</td>
<td></td>
</tr>
<tr>
<td>RbtK</td>
<td></td>
<td>535</td>
<td>60.0 (112);‡</td>
<td>58.850</td>
<td>d-Ribulose</td>
<td>Hexokinase</td>
<td></td>
</tr>
<tr>
<td>DalT</td>
<td>Transporter</td>
<td>425</td>
<td>–</td>
<td>46.750</td>
<td>d-Arabinitol</td>
<td>Ion symporter</td>
<td>86.6</td>
</tr>
<tr>
<td>RbtT</td>
<td></td>
<td>427</td>
<td>425§</td>
<td>46.970</td>
<td>Ribitol</td>
<td>Ion symporter</td>
<td></td>
</tr>
</tbody>
</table>

* Local similarities, e.g. for helix–turn–helix motifs or NAD- and ATP-binding motifs, are not considered.
† Purified monomer.
‡ Long and short dehydrogenases, according to Persson et al. (1991a, b), active as monomers or as tetramers, respectively.
§ Data from Heuel et al. (1997).
polymerase-dependent overexpression system (Heuel et al., 1997).

**DNA sequence analysis and protein comparison**

After complete sequencing of the *dal* and *rbt* genes, including the regulatory genes, sequence alignment and comparisons at the DNA and peptide level could be done. Due to previously unresolved DNA compressions, the published sequences had to be corrected in numerous places. These corrections, however, do not alter previous major conclusions. On the other hand, the comparisons revealed some new and unexpected results (Fig. 2 and Table 1).

**dalR and rbtR encoding repressors DalR and RbtR**

*dalR* and *rbtR* are transcribed convergently from promoters *attB* and *attP*, respectively (Fig. 2). The secondary *attB* site (nt 5777-5921), used by Hartley and co-workers (Neuberger & Hartley, 1979) to clone the *dal* and *rbt* genes onto phage λ, is located within the intergenic sequence of 209 bp. The RbtR repressor (270 aa), which recognizes the molecular inducer d-ribitol, belongs to the LacI/GalR family. It contains the typical amino-terminal helix-turn-helix motif (aa5-25) and other characteristic motifs in the central and carboxy-terminal parts (Weickert & Adhya, 1992). The DalR repressor (313 aa), on the other hand, binds d-arabinitol as the molecular inducer and belongs to the DeoR repressor family (van Rooijen & de Vos, 1990). Characteristic members of this group of DeoR (deoxyribose; about 32% identical residues), GatR (galactitol), GutR (d-glucitol) and SorC (l-sorbose metabolism) in enteric bacteria. All contain a common inducer-binding motif near the carboxy-terminal end and an amino-terminal helix-turn-helix motif (26-45). Neither *dalR* and *rbtR* at the DNA level, nor DalR and RbtR at the peptide level showed any significant similarity.

**dalD and rbtD encoding d-arabinitol (DalD) and ribitol dehydrogenase (RbtD)**

In both operons, the promoter-proximal genes *dalD* and *rbtD* encode pentose-specific and NAD-dependent dehydrogenases. Against expectation, however, neither *dalD* and *rbtD* nor the corresponding proteins showed any similarity which identified the gene products as members of the dehydrogenase family. Thus, DalD (455 aa) with affinity to d-arabinitol (*K_M* 24 mM) and d-ribitol (*K_M* 70 mM), but not to ribitol, xylitol and d-glucitol, belongs to the subfamily of long (≥450 aa) dehydrogenases (Persson et al., 1991a, b). These contain a typical NAD-binding (aa9-17) and substrate-binding (205-211) motif also conserved in the d-ribitol-1-phosphate dehydrogenase of, e.g. *E. coli* K-12. Long dehydrogenases are active as monomers as has also been shown for DalD (Hartley, 1984b). RbtD (248 aa), by contrast, belongs to the subfamily of short dehydrogenases and acts as a tetramer. Substrates are (in decreasing order) ribitol (*K_M* 11 mM), L-arabinitol (267 mM) and xylitol (1000 mM), but not d-arabinitol and d-mannitol (Hartley, 1984b). An NAD-binding motif (21-27) was found as expected.

**dalK and rbtK encoding d-xylulose (DalK) and d-ribulose kinase (RbtK)**

*dalK* and *rbtK* encode ATP-dependent kinases for d-xylulose and d-ribulose, respectively (Fig. 3). As for the dehydrogenase genes, *dalK* and *rbtK* did not show extended similarities. The d-xylulose kinase DalK (487 aa) is active as a dimer (about 110000 Da) and resembles the xylB-encoded d-xylulokinase from *K. pneumoniae* (484 aa; 52.9% identical residues). For comparison, the XylB kinases from *K. pneumoniae* and *E. coli* share 81.3% identical residues. Among the many pentoses, pentitols and hexitols, only d-xylulose (*K_M* 0.8 mM; *V_{max} 150 μmol min^{-1} (mg protein)^{-1}) is a substrate for DalK (Hartley, 1984b). It has been claimed that XylB accepts d-arabinitol and produces d-arabinitol 5-phosphate (Scangos & Reiner, 1979). The three kinases belong to the prokaryotic d-xylulokinase group of the hexokinase family as defined by Bork et al. (1993). Based on three-dimensional crystal structures, hexokinases comprise characteristic structures: in DalK...
Fig. 4. Mapping of the transcription initiation sites of dalR (a) and dalD (b) by primer extension analysis. Radiolabelled primers complementary to the 5' end of the non-coding strand of dalR (5' CCATCGCCGACGCCACCTTCTGATCCACCGG 3') or dalD (5' CCACGCCTGATGGCGCGGATGAAAAAGAACC 3') were used to direct cDNA synthesis by reverse transcription from total RNAs of cells of strain HB101 containing either pPHL101 (dalDKT) or pHPL102 (dalIk). The respective extended primers (+1) are shown in the right-hand lanes. The sequence of the complementary DNA is also given. This was derived from lanes A, C, G and T showing the sequencing ladders with primers labelled at 5' positions by T4 polynucleotide kinase. An asterisk indicates the initiation site, namely an A residue. 50 bp upstream of the putative ATG codon of dalR and 21 bp upstream of the putative ATG codon of dalD.

(Fig. 3) these correspond to two phosphate-binding sites (aa 2–22 and 251–266) and an adenosine binding site (388–417), all three involved in ATP binding, two connect or linker structures (230–248 and 421–441) and a putative substrate-binding site (68–86).

The kinase encoded by rbtK (535 aa) also acts as a dimer (about 112000 Da). Substrates are D-ribulose \( [K_M^\text{app} = 0.4 \text{ mM}] \) and D-glucurionate \( [V_{\text{max}} = 71 \text{ pmol min}^{-1} (\text{mg protein})^{-1}] \) and to a lesser extent D-arabinitol and ribitol \( [K_M^\text{app} = 140 \text{ and } 220 \text{ mM}] \), respectively but not D-xylulose (Hartley, 1984b). Based on sequence alignments, RbtK corresponds to a typical prokaryotic hexokinase, in particular to the group of t-ribulose-, t-glucurionate- and glycerol kinases (Bork et al., 1993). Curiously, its closest relative is the hydrophilic part of a large (715 aa) and allegedly membrane-bound protein from Saccharomyces cerevisiae (accession no. Z48785). RbtK shares 25–53 % identical residues with the kinases, the characteristic phosphate 1 and 2 sites (aa 8–28 and 290–305), the connect 1 and 2 motifs (269–287 and 468–488), and the adenosine-binding site (435–464), the latter preceded by a helix motif (411–423) (Fig. 3). The total DalK and RbtK molecules share 19–3 % identical residues.

Operons dal and rbt are not flanked by inverted repeats

It has been claimed repeatedly that dal and rbt genes are sandwiched between rather long (0.4–1.4 kbp) inverted repeats. It has been speculated that these could be the remnants of a metabolic transposon (Hartley, 1984b) and perhaps could be involved in genotypic exclusion between the atl and rlt genes and the gat genes in E. coli (Link & Reiner, 1982). We have shown recently (Heuel et al., 1997) that the dal and rbt operons of K. pneumoniae contain previously unidentified genes, dalT and rbtT that are located proximal to genes dalK and rbtK, respectively (Fig. 2). The genes encode D-arabinitol- and ribitol-specific transporters of unusually high similarity (86.6 % identical amino acids) that is reflected at the DNA level (82.7 % identical bases). The postulated inverted repeats correspond to parts of both genes (see IR in Fig. 2), but not to short inverted repeats as they are characteristic of transposable elements. The pronounced similarity starts exactly 6 bp before the ribosome-binding site of dalT and rbtT within intragenic sequences which otherwise did not show any similarity. It ends abruptly within a TTC codon for dalT and a TTT codon for rbtT located 42 and 48 bp before the corresponding stop codons, hence 14 and 16 aa before the end of the peptides, respectively. These residues are parts of the terminal cytoplasmic loop of the permeases. At 58 bp downstream from rbtT, dbnA begins (Fig. 2), a gene found in all atl rlt (dal rbt) and gat strains of E. coli and K. pneumoniae tested thus far (unpublished results). dalT on the other side is followed by an ORF not present on the chromosome of E. coli K-12.

Regulatory sequences involved in transcription control and size of the complete dal and rbt operons

Sequence alignments have revealed putative promoter, CRP-binding and ribosome-binding motifs as well as putative initiation and stop codons, DalD, DalK, RbtK and RbtD have been purified to homogeneity. Their amino-terminal sequences have been used to define putative ribosome-binding sites and initiation codons (references in Hartley, 1984b; Wu et al., 1985).

In the case of the dalRp and the dalDp promoters, the transcription initiation bases were determined directly by primer extension analysis as described in Methods and in the legend to Fig. 4. These are an A residue located 50 bp upstream of the dalR initiation codon and an A residue located 21 bp upstream of the dalD initiation codon. Promoter dalDp contains typical −10 (TACAGT) and −35 (TTATT) sequences, the latter preceded immediately by a CRP-binding consensus motif (TGTGA \( N_x \) GCTCT) with an imperfect symmetry (Fig. 5). Four short palindromes (AATTA) interspersed between these sequences are possibly involved in DalR binding as operators. Similar structures are not found in front of dalK and in dalRp. Consistent with these results it was observed that the dal operon in KAY2026 was inducible by external D-arabinitol as tested by DalT activity [1 and 46 nmol transport min\(^{-1}\) (mg protein\(^{-1}\)] for uninduced and induced cells, respectively and that deletion of dalR produced fully constitutive clones [16 nmol transport min\(^{-1}\) (mg protein\(^{-1}\)] for uninduced cells. Induction was repressed in the presence of glucose (Knott, 1982).
The ends of Fig. 6. for phage separated by 209 bp. This includes a secondary attachment site deduced peptide sequences and stop codons. indicated (asterisks).

dalT is followed by a short intergenic sequence (48 bp) and an ORF (orf1) of about 1 kbp length. orf1 did not show similarity to any sequence in the sequence databases. Deletion of orf1, as a HindIII-SphI fragment from pHLH101 (Fig. 2), caused no visible change in the Dal phenotype of transformants and preliminary Northern hybridization studies gave no indication for the presence of orf1-specific mRNA in pHLH101 transformants. These results seem to rule out orf1 as a member of the dalDKT operon.

Repressors DalR and RbtR showed no similarity. In accordance with this result, there was no sequence similarity between the intergenic sequences separating dalR and dalD, and rbtR and rbtD, respectively, that carry the corresponding repressor-binding sites. Putative −10 and −35 consensus sequences for promoters rbtDp and rbtRp have been postulated (Wu et al., 1985). Neither these nor a putative CRP-binding sequence(s) in front of rbtDp have been identified and characterized by direct tests. The expression of the rbt genes is, however, strongly reduced during growth on glucose (Neuberger & Hartley, 1979). Early attempts to follow rbtD-specific mRNA synthesis by identifying polysomes engaged in RbtD synthesis through RbtD-specific antibodies, indicated a dramatic shift in the mid-exponential phase of growth, i.e. a fivefold increase in RbtD activity paralleled by an increase in corresponding polysomes (Hartley, 1984b). It remains to be shown which mechanism underlies this pattern.

Conclusions

As stated above, the dal and rbt genes have been considered as models for the natural evolution of new operons and metabolic pathways by gene duplication and for metabolic transposons. The present results corroborate previous doubts voiced by Hartley (1984b) on this too simplistic view. Thus, dalR, D and K do not share extensive similarity with rbtR, D and K, respectively, and the corresponding repressors, dehydrogenases and kinases belong to different protein families. This argues against evolution by gene or operon duplication followed by diversification and argues for modular evolution ('gene swapping') in which the various genes have been assembled in a mosaic-like way into their present positions. The only area of extensive similarity covers dalT and rbtT. Similarity is so high (82.7 % identical bases and 86.6 % identical amino acid residues) over the entire length of the genes/peptides that it almost certainly indicates a recent origin from a common ancestor, i.e. true homology. It is these DNAs (Fig. 2) which have been considered previously as inverted repeats and the remnants of a putative metabolic transposon (Link & Reiner, 1982, 1983; Neuberger & Hartley, 1979).

The evolution of the catabolic pathways for D-arabinitol and ribitol must have been closely linked among the enteric bacteria. Thus, based on biochemical studies, both catabolic pathways are highly similar in Klebsiella (Fig. 1) and in E. coli C (Reiner, 1975), and based on genetic studies the operons are very similar in both organisms and are located at a similar place in the chromosome (Scangos & Reiner, 1979). We cloned the att and rtl genes from E. coli C and were able to show that they map in both organisms at an identical location,
that is between the attP2 site and the dhnA gene (Fig. 2). This location may explain why Rigby et al. (1976) were able to transduce the dal and rbt genes from K. pneumoniae to E. coli K-12 by means of phage P1. Partial sequencing of the att and rtl genes revealed the expected high similarity with the dal and rbt genes (about 82% identical bases) and the same operon structure (unpublished results). We were also able to confirm the presence of a secondary attachment site for phage λ located between the dal and rbt genes (Fig. 6), which allowed the original cloning of these genes into phage λ derivatives (Neuberger & Hartley, 1979). Finally, we were able to confirm the presence of chi-like recombination sites (GCCTGCC) within dalK (nt 1921–1927) and rbtK (7974–7980) (see also Fig. 2). These apparently became active in the selection of strains overproducing RbtD during continuous culture on xylitol and caused amplification of dalDR and rbtRD (Neuberger & Hartley, 1979). These results strongly indicate the presence of a natural mechanism for interspecies horizontal gene transfer of the pentitol-specific genes and enzymes within enteric bacteria.

ACKNOWLEDGEMENTS

We thank E. Placke for help in preparing the manuscript. For financial support through SFB171 (H.H. and S.T.) and a Graduierten-stipendium (A.S.-G.) we thank the Deutsche Forschungsgemeinschaft and the Verband der Chemischen Industrie for general financial support.

REFERENCES


Received 20 November 1997; revised 12 February 1998; accepted 16 February 1998.